Derandomizing Codes for the Binary Adversarial Wiretap Channel of Type II

Eric Ruzomberka*, Homa Nikbakht*, Christopher G. Brinton†, David J. Love† and H. Vincent Poor*

*Princeton University †Purdue University

Abstract—We revisit the binary adversarial wiretap channel (AWTC) of type II in which an active adversary can read a fraction r and flip a fraction p of codeword bits. The semantic-secrecy capacity of the AWTC II is partially known, where the best-known lower bound is non-constructive, proven via a random coding argument that uses a large number (that is exponential in blocklength n) of random bits to seed the random code. In this paper, we establish a new derandomization result in which we match the best-known lower bound of $1 - H_2(p) - r$ where $H_2(\cdot)$ is the binary entropy function via a random code that uses a small seed of only $O(n^2)$ bits. Our random code construction is a novel application of *pseudolinear codes* – a class of non-linear codes that have k-wise independent codewords when picked at random where k is a design parameter. As the key technical tool in our analysis, we provide a soft-covering lemma in the flavor of Goldfeld, Cuff and Permuter (Trans. Inf. Theory 2016) that holds for random codes with k-wise independent codewords.

I. INTRODUCTION

Consider a communication setting in which a sender Alice wishes to communicate a message to a receiver Bob by sending a sequence of bits over a noisy wiretap channel. The channel is controlled by an (active) adversary who can both read a fraction $r \in [0, 1]$ and flip a fraction $p \in [0, 1]$ of Alice's transmitted bits. In this setting, Alice's and Bob's communication goal under any adversarial strategy is two-fold:

1) (*Reliability*) Bob must decode Alice's message with small probability of error.
2) (*Secrecy*) The adversary must extract negligible information of the Alice's message via its observation of Alice's sequence.

Critically, we make no assumptions about the adversary's computational limitations, and thus, secrecy must be guaranteed in an information theoretic sense by "hiding" the message in the adversary's bit-limited observation. Furthermore, the adversary may choose the location of the bit reads and bit flips in an arbitrary manner using knowledge of Alice and Bob communication scheme. In the literature, the above setting is known as the binary adversarial wiretap channel of type II (denoted as (p, r)-AWTC II) [1], [2].

Much is known about the fundamental limits of communication over the (p, r)-AWTC II. Roughly defined, the secrecy capacity of the (p, r)-AWTC II is the largest rate at which

Alice and Bob can communicate while meeting the above goals under a given secrecy measure. The measure we focus on is semantic secrecy (SS) [3], [4], which is widely recognized as the cryptographic gold standard for evaluating secrecy [5]. The SS capacity, denoted $C(p, r)$, is partially known where the best-known lower bound [6] and upper bound [6], [7] are

$$\max\{1 - H_2(p) - r, 0\} \le C(p, r) \le 1 - H_2(p) - r - \min_{x \in [0,1]} f(x) \tag{1}$$

where $H_2(\cdot)$ is the binary entropy function and $f(x) = H_2((2p-1)x + 1 - p) - H_2(p) - rH_2(x)$. Note that the two bounds are close for small r and tight for $p = 0$.

While the limits of communication over the (p, r)-AWTC II are mostly understood, less is known on how to construct efficient codes to achieve these limits. The proof of the lower bound (1), as given in [6], is non-constructive and follows an ordinary random coding argument in which codewords are chosen uniformly and *independently* from space $\{0, 1\}^n$ where n is the blocklength of the code. As a tool for probabilistic constructions, the practical use of this random code distribution is limited. For example, to represent a code picked in this way, one would need to remember at least $n2^{Rn}$ random bits where R is the coding rate.[1] Thus, codes picked from a distribution with mutual independence property lack a succinct representation. Furthermore, the high degree of randomness used in the construction obscures insight into the structure of a good code. Without sufficient structure, efficient encoding and decoding algorithms are likely to be elusive.

In this paper, we work towards an efficient code construction for the (p, r)-AWTC II by partially derandomizing the random code used in [6] to establish the lower bound (1). We do so by relaxing the requirement that codewords be mutually independent and consider random codes with k-wise independent codewords for some positive integer $k << n$. We show that random codes under this weaker notation of independence can achieve the lower bound (1) for some parameter k large enough but constant in n. As a result, these codes have both a more succinct representation and additional structure compared to random codes with mutually independent codewords.

The approach we take is the following. We focus on a class of non-linear codes known as *pseudolinear codes* (precisely defined in Section II-C), which was initially proposed by

[1]Additional random (seed) bits are needed if one considers codes with private randomness at the encoder (i.e., stochastic codes).

Guruswami and Indyk [8] outside of the AWTC setting. In the AWTC setting, pseudolinear codes have a number of nice properties, including succinct representations (i.e., $O(kn^2)$ bits), efficient encoding algorithms, some linear structure, and k-wise independent codewords when chosen at random for a designable parameter k. We initiate the study of pseudolinear codes for achieving both secrecy and reliability in the wiretap setting. As our main result, we show that random pseudolinear codes achieve the best-known SS capacity lower bound (1). Conversely, we show that non-linear codes are *necessary* to achieve this lower bound for some values of p and r. To prove our main result, we provide a new lemma on the soft-covering phenomenon [9], [10] under random coding with k-wise independent codewords.

II. Preliminaries, Results & Related Work

A. Notation

Unless stated otherwise, we denote random variables in uppercase font (e.g., X), realizations of random variables in lowercase font (e.g., x), and sequences in bold font (e.g., $\boldsymbol{X}$, $\boldsymbol{x}$). An exception to the above rules occurs when we denote codes: we denote random codes with script typeset (e.g., $\mathscr{C}$) and realizations of random codes with calligraphic typeset (e.g., $\mathcal{C}$). We denote the set of all possible distributions over a set $\mathcal{X}$ as $\mathcal{P}(\mathcal{X})$, and denote the uniform distribution over $\mathcal{X}$ as $\mathrm{Unif}(\mathcal{X})$. We denote that X is distributed as $P \in \mathcal{P}(\mathcal{X})$ by writing $X \sim P$. For PMFs P and Q such that $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$ (absolute continuity), the relative entropy of P and Q is $D(P||Q) \triangleq \sum_{x \in \mathrm{supp}(P)} P(x) \log_2 \frac{P(x)}{Q(x)}$. For $\alpha > 0$ and $\alpha \neq 1$, the Rényi divergence of order α is $D_\alpha(P||Q) \triangleq \frac{1}{\alpha-1} \log_2 \sum_{x \in \mathrm{supp}(P)} P(x)(\frac{P(x)}{Q(x)})^{\alpha-1}$. Define the special case $D_1(P||Q) \triangleq \lim_{\alpha \to 1} D_\alpha(P||Q) = D(P||Q)$. For an event $\mathcal{A}$, we let $\mathbb{1}\{\mathcal{A}\}$ denote the indicator of $\mathcal{A}$.

B. Setup

Code: A (binary) code $\mathcal{C}_n$ of blocklength n is a subset of $\{0,1\}^n$. We will associate a code $\mathcal{C}_n$ with an encoding function $\boldsymbol{x}(\cdot)$, which performs a mapping from the message space $\mathcal{M}$ to codewords in $\{0,1\}^n$. As is common for wiretap codes, we will consider *stochastic encoding* in which $\boldsymbol{x}$ takes as argument a private random key $w \in \mathcal{W}$ that is known only to Alice. Specifically, for a message rate $R = \frac{\log_2 |\mathcal{M}|}{n}$ and a (private) key rate $R' = \frac{\log_2 |\mathcal{W}|}{n}$, an $[n, Rn, R'n]$ code $\mathcal{C}_n$ is a set

$$\mathcal{C}_n = \{\boldsymbol{x}(m,w) : (m,w) \in \mathcal{M} \times \mathcal{W}\}$$

where we refer to $\boldsymbol{x}(w,m)$ as the (n-bit) codeword corresponding to message m and key w. In turn, a *family* of codes is a sequence $\{\mathcal{C}_n\}_{n=1}^\infty$ where for each $n \geq 1$, $\mathcal{C}_n$ is an $[n, Rn, R'n]$ code.

Encoding/Decoding: For an $[n, Rn, R'n]$ code $\mathcal{C}_n$, probability mass function (PMF) $P_M \in \mathcal{P}(\mathcal{M})$, a message $M \sim P_M$ and a private key $W \sim \mathrm{Unif}(\mathcal{W})$ where M and W are independent, Alice encodes M into a codeword $\boldsymbol{x}(M,W)$ and transmits it over the channel. Subsequently, Bob receives a corrupted version of the codeword and performs decoding by choosing a message estimate $\widehat{M} \in \mathcal{M}$. We say that a decoding error occurs if $\widehat{M} \neq M$.

The AWTC II: For a read fraction $r \in [0,1]$ and an error fraction $p \in [0, 1/2]$, the adversary can observe rn bits and flip up to pn bits of $\boldsymbol{x}(M,W)$. The location of the read bits are indexed by a coordinate set $\mathcal{S}$, which the adversary can choose from the set $\mathscr{S}$ consisting of all subsets of $[n]$ of size rn. In turn, the adversary observes $\boldsymbol{Z} = \boldsymbol{x}(M,W,\mathcal{S})$ where $\boldsymbol{x}(M,W,\mathcal{S})$ denotes the rn bits of $\boldsymbol{x}(M,W)$ indexed at $\mathcal{S}$, and subsequently, chooses the location of the bit flips. We emphasize that the location of the bit flips need not coincide with $\mathcal{S}$. In general, the adversary can randomize its above choices by choosing a distribution on $\mathcal{S}$ that can depend on the code, as well as a distribution on the bit flip locations that can depend on both the code and the observation $\boldsymbol{Z}$.

Secrecy: Define the semantic leakage as

$$\mathrm{Sem}(\mathcal{C}_n) = \max_{P_M \in P(\mathcal{M}), \mathcal{S} \in \mathscr{S}} I_\mathcal{S}(M; \boldsymbol{Z}) \qquad (2)$$

where $I_\mathcal{S}(M;Z)$ denotes the mutual information between $M \sim P_M$ and $\boldsymbol{Z} = \boldsymbol{x}(M,W,\mathcal{S})$. In turn, a family of codes $\{\mathcal{C}_n\}_{n=1}^\infty$ is said to be *semantically-secret* if $\mathrm{Sem}(\mathcal{C}_n) = 2^{-\Omega(n)}$. We remark that this mutual-information based notation of SS is shown in [5] to be (asymptotically) equivalent to the operational definition of SS given in [3], [4]. Further, SS is a stronger notation of secrecy than strong secrecy.[2]

Reliability: The (maximum) probability of decoding error is defined as

$$P_{\mathrm{error}}^{\max}(\mathcal{C}_n) = \max_{m \in \mathcal{M}} \mathbb{P}\left(\widehat{M} \neq m | M = m\right)$$

where the probability is taken w.r.t. the distribution of Alice's key and the *worst-case* distribution of the adversary's bit read/flip locations. A family of codes $\{\mathcal{C}_n\}_{n=1}^\infty$ is said to be *reliable* if for any $\delta > 0$, $P_{\mathrm{error}}(\mathcal{C}_n) \leq \delta$ for large enough n.

SS Capacity: The rate $R > 0$ is said to be achievable over the (p,r)-AWTC II if there exists a family of codes $\{\mathcal{C}_n\}_{n=1}^\infty$ (where for each n, $\mathcal{C}_n$ is an $[n, Rn, R'n]$ code for some $R' \geq 0$) that is both semantically-secret and reliable. The SS capacity $C(p,r)$ is the supremum of rates achievable over the (p,r)-AWTC II.

C. Results

Our first result is on the necessity of non-linear codes for achieving the SS capacity. We say that a $[n, Rn, R'n]$ code $\mathcal{C}_n$ is *linear*[3] if there exists a generator matrix $G \in \{0,1\}^{(R+R')n \times n}$ such that the codeword corresponding to any message $m \in \mathcal{M} \triangleq \{0,1\}^{Rn}$ and key $w \in \mathcal{W} \triangleq \{0,1\}^{R'n}$ is $\boldsymbol{x}(m,w) = \begin{bmatrix} m & w \end{bmatrix} G$. A corollary of the following Theorem is that for any $r \in (0,1]$ and $p = 0$ (i.e., the channel to Bob is noiseless), linear codes cannot achieve SS capacity $C(0,r) = \max\{1 - r, 0\}$.

---

[2] A family of codes is said to achieve strong secrecy if $\lim_{n \to \infty} \max_{\mathcal{S} \in \mathscr{S}} I_\mathcal{S}(M; \boldsymbol{Z}) = 0$ where the message distribution is fixed s.t. $P_M \sim \mathrm{Unif}(\mathcal{M})$.

[3] Examples of linear codes in the wiretap setting include Ozarow's and Wyner's linear coset coding scheme [1] and some polar code and LDPC code based schemes (e.g., [11]).
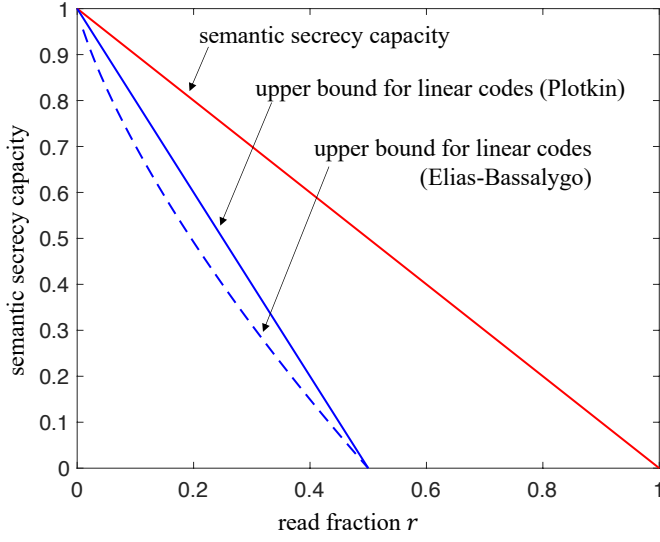
Fig. 1: Bounds on linear code performance compared to the semantic secrecy capacity of the (p, r)-AWTC II when the channel from Alice to Bob is noiseless (i.e., $p = 0$).

Theorem 1. *Let $p = 0$, $r \in (0, 1]$, $R > \max\{1 - 2r, 0\}$ and $R' \in [0, 1 - R]$. For large enough n, every linear $[n, Rn, R'n]$ code $\mathcal{C}_n$ has either semantic leakage $\mathrm{Sem}(\mathcal{C}_n) \geq 1$ or probability of error $P_{\mathrm{error}}(\mathcal{C}_n) \geq 1/2$ over the $(0, r)$-AWTC II.*

Remark 1. *Theorem 1 can be extended to non-zero values of p. In particular, together with the lower bound (1), Theorem 1 implies that linear codes cannot achieve $C(p, r)$ for either any $p \in [0, 1/2)$ and $r \in (0, 1/2]$ such that $H_2(p) < r$, or any $p \in [0, 1/2]$ and $r \in [1/2, 1]$, except for the trivial case when $C(p, r)$ is 0.*

A proof of Theorem 1 is given in Section III, which involves a specific construction of the adversary's coordinates $\mathcal{S}$ together with the Plotkin bound to upper bound the minimum distance of a code. We remark that tighter distance bounds can be used in place of the Plotkin bound. For instance, if one uses the Elias-Bassalygo bound [12]–[14], the rate lower bound in Theorem 1 can be tightened to $R > \max\{1 - H\left(\frac{1 - \sqrt{1 - 2r}}{2}\right), 0\}$. All bounds discussed thus far are plotted in Fig. 1.

In light of Theorem 1, non-linear codes must be considered to achieve the lower bound (1) for at least some values of $p \in [0, 1/2]$ and $r \in [0, 1]$. We turn now to non-linear codes.

Definition 1 (Pseudolinear Code)**.** *For $R \in (0, 1]$, $R' \in [0, 1 - R]$ and positive integers n and k, let H be the parity check matrix of any binary linear code with the following parameters:*

- *blocklength $2^{(R+R')n} - 1$*
- *dimension $2^{(R+R')n} - 1 - \ell$ for some $\ell = O(k(R + R')n)$*
- *minimum distance at least $k + 1$.*

An $[n, Rn, R'n, k]$ psuedolinear code $\mathcal{C}_n$ is any $[n, Rn, R'n]$

code that satisfies the following two step encoding process. First, a message-key pair $(m, w) \in \mathcal{M} \times \mathcal{W}$ is mapped to the row of H^T indexed by (m, w), which we denote as $\boldsymbol{h}(m, w)$.[4] Second, $\boldsymbol{h}(m, w)$ is linearly mapped to an n-bit codeword by some "generator" matrix $G \in \{0, 1\}^{\ell \times n}$, i.e.,

$$\boldsymbol{x}(m, w) = \boldsymbol{h}(m, w)G.$$

Thus, the non-linearity of $\mathcal{C}_n$ is confined to the first stage of encoding.

Towards the goal of derandomizing the random code of [6], pseudolinear codes have the following three attractive properties [8].[5] First, a pseudolinear code has a succinct representation as only $\ell n = O(k(R + R')n^2)$ bits are needed to describe the generator matrix. Second, encoding is computationally efficient if $\boldsymbol{h}(m, w)$ can be obtained in time polynomial in n for each $(m, w) \in \mathcal{M} \times \mathcal{W}$. For instance, we can let H be the parity check matrix of a binary Bose–Chaudhuri–Hocquenghem (BCH) code of design distance $k + 1$, in which case H has an explicit representation and $\boldsymbol{h}(m, w)$ can be efficiently obtained by computing powers of a primitive $(2^{(R+R')n} - 1)$-th root of unity from the extension field $\mathrm{GF}(2^{(R+R')n})$, e.g., see [16].

Third, if we consider a *random* pseudolinear code by choosing the generator matrix G at random while fixing the parity check matrix H, then the codewords of the random code are uniformly distributed in $\{0, 1\}^n$ and k-wise independent, i.e., any subset of codewords of size k are mutually independent.[6] This final property is the key to showing that pseudolinear codes achieve the best-known lower bound of $C(p, r)$.

Theorem 2. *Let $p \in [0, 1/2]$ and $r \in [0, 1]$ such that $1 - H_2(p) - r$ is positive. For any $R < 1 - H_2(p) - r$ and for large enough (but fixed) k, there exists a family pseudolinear codes $\{\mathcal{C}_n\}_{n=1}^{\infty}$ (where for $n \geq 1$, $\mathcal{C}_n$ is an $[n, Rn, R'n, k]$ pseudolinear code for some $R' \geq 0$) that is both reliable and semantically-secret.*

A proof of Theorem 2 is provided in Section V. The key technical tool in the proof is a new version of Wyner's soft-covering lemma which holds for codes with k-wise independent codeword. However, our version differs significantly from Wyner's [9, Theorem 6.3], which we state and prove in Section IV.

Our version is closest to (and proved similarly to) the soft-covering lemma of Goldfeld, Cuff and Permuter [10], which roughly states that if the key rate R' is larger than the mutual information between Alice's channel input and the adversary's observation, then a random code with mutually independent codewords satisfies an exponential number of secrecy constraints with probability at least $1 - 2^{-2^{\Omega(n)}}$. Here, the double-exponential probability bound is important as it

---

[4]To account for the message-key pair $(0, 0)$, we define $\boldsymbol{h}(0, 0)$ to be the all zeros vector.

[5]See [15] for further discussion of pseudolinear codes.

[6]In contrast, random linear codes have codewords that are pair-wise (i.e., 2-wise) independent in non-trivial cases.

allows one to take a union bound over an exponential number of events. Our version of the lemma states that when we restrict the random code to a k-wise independent distribution, the same constraints hold with probability at least $1-2^{-k\Omega(n)}$. Critically, while our probability bound tends to 1 more slowly than double-exponentially, it remains fast enough to take a union bound over an exponential number of events when k is large enough.

D. Related Work

Linear Codes and Semantic-Secrecy: Recall that Theorem 1 states that linear codes cannot achieve the SS capacity for the (noiseless) $(0, r)$-AWTC II for any $r \in (0, 1]$. Prior to this work, some special classes of linear codes were known to not achieve the SS capacity. In particular, Ozarow's and Wyner's linear coset coding scheme [1] does not achieve SS capacity of the $(0, r)$-AWTC II for any $r \in (0, 1]$. We provide a proof of this result in Appendix A. We remark that the necessity of non-linear codes for achieving the secrecy capacity is a product of the *joint consideration* of the semantic secrecy metric and the type II property of the wiretap channel. In contrast, linear codes are sufficient to achieve the *weak* secrecy capacity over the noiseless WTC II [1]. Furthermore, linear codes are sufficient to achieve both the weak and strong secrecy capacity of the noisy (but non-adversarial) WTC I [11].

Code Constructions: Explicit (and efficient) constructions that achieve the best known lower bound of the (p, r)-AWTC II are not known in general, except for the special cases of $p = 0$ [17], [18] and $r = 0$ [19]. In the general case, one promising approach is use modular constructions, which combine an existing error-control code with an invertible extractor [5], [18], [20] or algebraic manipulation detection code [2]. However, constructing binary error-control codes that are both efficiently encodable/decodable and achieve the (reliability) capacity of the (p, r)-AWTC is an open problem. In contrast to the above modular constructions, pseudolinear codes offer a non-modular approach. Recently, random (and thus non-explicit) pseudolinear codes were shown to achieve the (reliability) capacity of the (p, r)-AWTC II [21].

III. PROOF OF THEOREM 1

Notation: For message rate $R > 0$, key rate $R' \in [0, 1 - R]$, and blocklength $n \geq 1$ define $\mathcal{M} \triangleq \{0, 1\}^{Rn}$ and $\mathcal{W} \triangleq \{0, 1\}^{R'n}$. For an $[n, Rn, R'n]$ linear code $\mathcal{C}_n$, let G denote the $(R + R')n \times n$ generator matrix of $\mathcal{C}_n$, which can be partitioned such that $G = \begin{bmatrix} G_M \\ G_W \end{bmatrix}$ where $G_M \in \{0, 1\}^{Rn \times n}$ and $G_W \in \{0, 1\}^{R'n \times n}$. In turn, the codeword corresponding to message $m \in \mathcal{M}$ and key $w \in \mathcal{W}$ is $\boldsymbol{x}(m, w) = mG_M + wG_W$. For a coordinate set $\mathcal{S} \in \mathscr{S}$, let the matrices $G_M(\mathcal{S})$ and $G_W(\mathcal{S})$ denote the columns of G_M and G_W indexed by $\mathcal{S}$, respectively. Using this notation, if Alice transmits codeword $\boldsymbol{x}(m, w)$ then the adversary observes $\boldsymbol{z} = mG_M(\mathcal{S}) + wG_W(\mathcal{S})$.

Preliminaries: Let $\mathcal{C}_n$ be an $[n, Rn, R'n]$ linear code with generator matrix G. We make the following assumption.

Assumption 1. *Without loss of generality (w.l.o.g.), we assume that G is full rank, i.e.,* $\mathrm{rank}(G) = (R + R')n$.

The claim being w.l.o.g. is roughly as follows: if G is not full rank, then either $P_{\mathrm{error}}^{\max}(\mathcal{C}_n) \geq 1/2$ or both $\mathcal{W}$ and G can be replaced with a smaller key set and full rank generator matrix, respectively, without changing the code. A detailed discussion is provided in Appendix B. We remark that following Assumption 1, we have that $\mathrm{rank}(G_M) = Rn$ and $\mathrm{rank}(G_W) = R'n$.

Before proving the converse result (Theorem 1), we state a few preliminary results relating the semantic leakage to the rank of $G_M(\mathcal{S})$ and $G_W(\mathcal{S})$ for $\mathcal{S} \in \mathscr{S}$. For a code $\mathcal{C}_n$ and coordinate set $\mathcal{S} \in \mathscr{Z}$, we denote the mutual information between M and $\boldsymbol{Z}$ as $I_\mathcal{S}(M; \boldsymbol{Z})$ (where the dependency on $\mathcal{C}_n$ is implied).

Lemma 1. *For $\mathcal{S} \in \mathscr{S}$ and M uniformly distributed over $\mathcal{M}$,*

$$I_\mathcal{S}(M; \boldsymbol{Z}) = \mathrm{rank}\left(G(S)\right) - \mathrm{rank}\left(G_W(\mathcal{S})\right).$$

Proof of Lemma 1. Let $\mathcal{S} \in \mathscr{S}$. We first characterize the joint PMF of M, W and $\boldsymbol{Z}$, which we denote as $P_{M,W,\boldsymbol{Z}}$. We drop the subscripts from the PMF $P_{M,W,\boldsymbol{Z}}$ and its marginal PMFs when the meaning is clear from the use of the realization variables m, w and $\boldsymbol{z}$.

For $\boldsymbol{z} \in \{0, 1\}^{rn}$ and $m \in \mathcal{M}$, we have that

$$P(\boldsymbol{z}|m) = \sum_{w \in \mathcal{W}} P(\boldsymbol{z}, w|m) \stackrel{\text{(a)}}{=} \sum_{w \in \mathcal{W}} P(\boldsymbol{z}|m, w)P(w)$$
$$\stackrel{\text{(b)}}{=} T_{m,\boldsymbol{z}} 2^{-R'n} \tag{3}$$

where (a) follows from the independence of M and W, (b) follows from $W \sim \mathrm{Unif}(\mathcal{W})$, and where $T_{m,\boldsymbol{z}} \triangleq \sum_{w \in \mathcal{W}} \mathbb{1}\{\boldsymbol{z} = mG_M(\mathcal{S}) + wG_W(\mathcal{S})\}$.

To simplify (3), define

$$\mathcal{T} \triangleq \{(m', \boldsymbol{z}') \in \mathcal{M} \times \{0, 1\}^{rn} : T_{m',\boldsymbol{z}'} \geq 1\}$$

and suppose that $(m, \boldsymbol{z}) \in \mathcal{T}$. By definition, there exists an $w \in \mathcal{W}$ such that $wG_W(\mathcal{S}) = mG_M(\mathcal{S}) + \boldsymbol{z}$. In turn, since the mapping $G_W(\mathcal{S}) : \mathcal{W} \to \{0, 1\}^{rn}$ is a linear transformation, there must be $2^{\mathrm{nullity}(G_W(\mathcal{S}))}$ number of $w \in \mathcal{W}$ such that $wG_W(\mathcal{S}) = mG_M(\mathcal{S}) + \boldsymbol{z}$ where $\mathrm{nullity}(G_W(\mathcal{S}))$ is the dimension of the null space of $G_W(\mathcal{S})$. By the rank-nullity theorem [22, Theorem 2], $2^{\mathrm{nullity}(G_W(\mathcal{S}))} = 2^{\dim(\mathcal{W}) - \mathrm{rank}(G_W(\mathcal{S}))} = 2^{R'n - \mathrm{rank}(G_W(\mathcal{S}))}$. In turn,

$$T_{m,\boldsymbol{z}} = \begin{cases} 2^{R'n - \mathrm{rank}(G_W(\mathcal{S}))}, & (m, \boldsymbol{z}) \in \mathcal{T} \\ 0, & (m, \boldsymbol{z}) \notin \mathcal{T}, \end{cases}$$

and in turn, following (3),

$$P(\boldsymbol{z}|m) = \begin{cases} 2^{-\mathrm{rank}(G_W(\mathcal{S}))}, & (m, \boldsymbol{z}) \in \mathcal{T} \\ 0, & (m, \boldsymbol{z}) \notin \mathcal{T}. \end{cases} \tag{4}$$

Repeating the above approach for the PMF of $\boldsymbol{Z}$, one can show using the assumption that m is uniformly distributed over $\mathcal{M} = \{0,1\}^{Rn}$ that

$$
P(\boldsymbol{z}) = \begin{cases} 2^{-\text{rank}(G(\mathcal{S}))}, & \exists m \in \mathcal{M} \text{ s.t. } (m, \boldsymbol{z}) \in \mathcal{T} \\ 0, & \forall m \in \mathcal{M}, \ (m, \boldsymbol{z}) \notin \mathcal{T}. \end{cases} \tag{5}
$$

Using the above PMFs, we evaluate the mutual information between M and $\boldsymbol{Z}$:

$$
\begin{aligned}
I_{\mathcal{S}}(M; \boldsymbol{Z}) &\triangleq \sum_{m \in \mathcal{M}} \sum_{\boldsymbol{z} \in \{0,1\}^{rn}} P(m, \boldsymbol{z}) \log_2 \frac{P(\boldsymbol{z}|m)}{P(\boldsymbol{z})} \\
&\overset{(c)}{=} \sum_{(m,\boldsymbol{z}) \in \mathcal{T}} P(m, \boldsymbol{z}) \log_2 2^{\text{rank}(G(\mathcal{S})) - \text{rank}(G_W(\mathcal{S}))} \\
&\overset{(d)}{=} \text{rank}(G(\mathcal{S})) - \text{rank}(G_W(\mathcal{S})).
\end{aligned}
$$

where (c) follows from (4), (5), and $P(m, \boldsymbol{z}) = 0 \ \forall (m, \boldsymbol{z}) \notin \mathcal{T}$, and (d) follows from $\sum_{(m,\boldsymbol{z}) \in \mathcal{T}} P(m, \boldsymbol{z}) = 1$. ∎

Corollary 1.1. *If $R' + R \le r$, then $\lim_{n \to \infty} \text{Sem}(\mathcal{C}_n) = \infty$.*

Proof of Corollary 1.1. Suppose that M is uniformly distributed and that $R + R' \le r$. Recall that G has rank $(R + R')n$ (c.f. Assumption 1). Since $R + R' \le r$, there exists a $\mathcal{S} \in \mathscr{S}$ such that $\text{rank}(G(\mathcal{S})) = \text{rank}(G) = (R + R')n$. Let $\mathcal{S}$ be this coordinate set. It follows that $\text{rank}(G_W(\mathcal{S})) = R'n$, and in turn, $I_{\mathcal{S}}(M; \boldsymbol{Z}) = Rn$ following Lemma 1. In conclusion, $\text{Sem}(\mathcal{C}_n) \ge Rn$. ∎

For the converse analysis, we will need the following version of the Plotkin bound [23].

Lemma 2 (Extended Plotkin bound [24])**.** *Suppose that Ψ is an $[n, Rn]$ code (not necessarily linear) with minimum distance $d_{\min} \in (0, n/2]$. Then for $\delta \triangleq d_{\min}/n$,*

$$
R \le 1 - 2\delta + o(1)
$$

where the $o(1)$ term tends to 0 as n tends to infinity.

Converse (Proof of Theorem 1) Setup: Set $p = 0$ and let $r \in [0, 1]$. For any $\epsilon > 0$, let $R = \max\{1 - 2r, 0\} + \epsilon$ and let $R' \in [0, 1 - R]$ such that $R + R' > r$ (c.f. Corollary 1.1). In turn, we let $\mathcal{C}_n$ be an $[n, Rn, R'n]$ linear code with generator matrix G. W.l.o.g., we assume that G is full rank (c.f. Assumption 1).

Converse Attack: The adversary orchestrates it attack in two steps. First, the adversary chooses an index set $\mathcal{V} \subseteq [n]$ of size $(R + R')n$ such that all columns of $G(\mathcal{V})$ are linearly independent. Note that such a set exists following our assumption that G is rank $(R + R')n$. Second, the adversary chooses a coordinate set $\mathcal{S}^* \in \mathscr{S}$ to be a subset of $\mathcal{V}$ that minimizes the rank of $G_W(\mathcal{S}^*)$. Once Alice transmits her codeword $\boldsymbol{x}(M, W)$, the adversary reads the codeword bits $\boldsymbol{Z} = \boldsymbol{x}(M, W, \mathcal{S}^*)$ corresponding to the coordinates $\mathcal{S}^*$ with corresponding mutual information $I_{\mathcal{S}^*}(M; \boldsymbol{Z})$.

Converse Analysis: The goal of the converse analysis is to show that $I_{\mathcal{S}^*}(M; \boldsymbol{Z}) \ge 1$. We remark that $\mathcal{S}^*$ is a strict subset of $\mathcal{V}$ following the inequality $r < R + R'$. This fact

together with the fact that all $|\mathcal{V}|$ column of $G(\mathcal{V})$ are linearly independent implies that the rank of $G(\mathcal{S}^*)$ is rn. In turn, following Lemma 1,

$$
I_{\mathcal{S}^*}(M; \boldsymbol{Z}) = rn - \text{rank}(G_W(\mathcal{S}^*)). \tag{6}
$$

In the converse analysis, we show that $rn - \text{rank}(G_W(\mathcal{S}^*)) \ge 1$.

We proceed with the following dual code perspective. Consider $G_W(\mathcal{V})$ as the $R'n \times (R + R')n$ generator matrix of some $[(R + R')n, R'n]$ linear code Ψ. In turn, let $G_W^\perp(\mathcal{V})$ denote the $Rn \times (R + R')n$ generator matrix of the $[(R + R')n, Rn]$ dual code $\Psi^\perp$ of Ψ. By definition, $G_W(\mathcal{V})$ is the parity check matrix corresponding to the generator matrix $G_W^\perp(\mathcal{V})$. Let $d_{\min}^\perp$ denote the minimum distance of $\Psi^\perp$. By the definition of the parity check matrix (e.g., see [16]), there exists $d_{\min}^\perp$ linearly dependent columns of the parity check matrix $G_W(\mathcal{V})$. Hence, if $d_{\min}^\perp \le rn$, then the adversary's choice of $\mathcal{S}^*$ contains the indices of these $d_{\min}^\perp$ linearly dependent columns of G_W, i.e, the rank of $G_W(\mathcal{S}^*)$ is bounded above by $rn - 1$. In turn, $I_{\mathcal{S}^*}(M; Z) \ge 1$ via (6). To complete the proof, we show that $d_{\min}^\perp \le rn$.

Applying the Plotkin bound (Lemma 2) to the dual code $\Psi^\perp$, we have that

$$
\frac{R}{R + R'} \le 1 - 2\delta^\perp + o(1) \tag{7}
$$

for the distance parameter $\delta^\perp \triangleq \frac{d_{\min}^\perp}{(R+R')n}$ and where the $o(1)$ term tends to 0 as n tends to infinity. In turn, for large enough n,

$$
\begin{aligned}
d_{\min}^\perp &\overset{(d)}{\le} \frac{R'n}{2} + o(n) \\
&\overset{(e)}{\le} \frac{2r - \epsilon}{2} + o(n) \\
&\overset{(f)}{<} rn
\end{aligned}
$$

where (d) follows from a rearrangement of (7), (e) follows from the setting of rate $R = \max\{1 - 2r, 0\} + \epsilon$ and the trivial inequalities $R + R' \le 1$ and $\max\{1 - 2r, 0\} \ge 1 - 2r$, and (f) follows for large enough n. In conclusion, for large enough n, $I_{\mathcal{S}^*}(M; \boldsymbol{Z}) \ge 1$ and thus $\text{Sem}(\mathcal{C}_n) \ge 1$.

IV. A Soft-Covering Lemma for k-wise Independent Codewords

Notation: In this section only, we consider a more general code model than that introduced in Section II-B. For an alphabet $\mathcal{U}$ which is not necessarily binary, a blocklength n and a (private) key rate $R' > 0$, we define an $[n, R'n]$ code $\mathcal{C}_n$ as a subset of $\mathcal{U}^n$ of size $|\mathcal{C}_n| = 2^{R'n}$. We will often describe $\mathcal{C}_n$ by its set of codewords $\{\boldsymbol{u}(w, \mathcal{C}_n)\}_{w \in \mathcal{W}}$ for a key set $\mathcal{W} = [2^{R'n}]$.

We introduce the soft-covering problem, depicted in Fig. 2. The problem setup is as follows. For a blocklength $n \ge 1$, let $\mathcal{C}_n = \{\boldsymbol{u}(w, \mathcal{C}_n)\}_{w \in \mathcal{W}}$ be an $[n, R'n]$ code. Given a finite input alphabet $\mathcal{U}$, an input distribution Q_U, a finite output alphabet $\mathcal{V}$ and channel $Q_{V|U}$, consider the PMFs induced
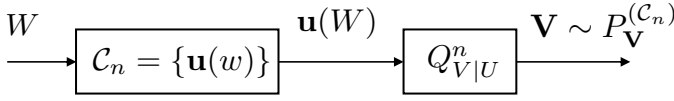
$$W \longrightarrow \boxed{\mathcal{C}_n = \{\mathbf{u}(w)\}} \xrightarrow{\mathbf{u}(W)} \boxed{Q^n_{V|U}} \xrightarrow{\mathbf{V} \sim P^{(\mathcal{C}_n)}_{\mathbf{V}}}$$

Fig. 2: The soft-covering problem: the goal is to design the code $\mathcal{C}_n$ to make $P^{(\mathcal{C}_n)}_{\mathbf{V}} \approx Q^n_V$.

on the output sequence $\mathbf{V} \in \mathcal{V}^n$ when an input sequence $\mathbf{U} \in \mathcal{U}^n$ is sent through the n-shot memoryless channel $Q^n_{V|U}$: for $\mathbf{v} \in \mathcal{V}^n$,

1) The PMF of $\mathbf{V}$ when $\mathbf{U}$ is drawn randomly from Q^n_U, i.e.,

$$Q_{\mathbf{V}}(\mathbf{v}) = Q^n_V(\mathbf{v}) = \sum_{\mathbf{u} \in \mathcal{U}} Q^n_{V|U}(\mathbf{v}|\mathbf{u}) Q^n_U(\mathbf{u}).$$

2) The PMF of $\mathbf{V}$ when $\mathbf{U}$ is the codeword $\mathbf{u}(W, \mathcal{C}_n)$ for $W \sim \mathrm{Unif}(\mathcal{W})$, i.e.,

$$P^{(\mathcal{C}_n)}_{\mathbf{V}}(\mathbf{v}) \triangleq \sum_{w \in \mathcal{W}} Q^n_{V|U}(\mathbf{v}|\mathbf{u}(w, \mathcal{C}_n)) 2^{-Rn}. \tag{8}$$

The soft-covering problem asks how to design a code $\mathcal{C}_n$ such that the induced PMF $\mathcal{P}^{(\mathcal{C}_n)}_{\mathbf{V}}$ is approximately Q^n_V in the limit as n tends to infinity. The following lemma states that if $R' > I(U; V)$, then for any integer k large enough a *random* $[n, R'n]$ code $\mathscr{C}_n$ with k-wise independent codewords each drawn from distribution Q^n_U results in $P^{(\mathscr{C}_n)}_{\mathbf{V}} \approx Q^n_V$ for large enough n. Recall that we denote random codes with script typeface (e.g., $\mathscr{C}_n$) and we denote realizations of random codes with calligraphic typeface (e.g., $\mathcal{C}_n$).

Lemma 3 (Soft-covering lemma for k-wise independent codewords). *Suppose that the random code $\mathscr{C}_n$ has k-wise independent codewords for some even integer $k \geq 4$, each drawn from a PMF Q^n_U for finite $\mathcal{U}$. Let $Q_{V|U}$ be any conditional PMF where $|\mathcal{V}|$ is finite and let $R' > I(U; V)$. There exists some $\gamma_0 > 0$ and $\gamma_1 > 0$ that depend only on R' and $I(U; V)$ such that for large enough n*

$$\mathbb{P}_{\mathscr{C}_n}\left(D\left(P^{(\mathscr{C}_n)}_{\mathbf{V}} \middle\| Q^n_V \right) > 2^{-\gamma_1 n} \right) \leq 2^{(-k\gamma_0 + \log_2 |\mathcal{V}|)n}$$

where we recall that D is the relative entropy.

A. Overview of Proof of Lemma 3

Setup: Let the blocklength $n \geq 1$ and key rate $R' > I(U; V)$, and let k be a positive integer that will be set later. In turn, let $\mathscr{C}_n$ be a random $[n, R'n]$ code drawn from any distribution that has k-wise independent codewords each with marginal PMF Q^n_U.

The proof of Lemma 3 follows a two step approach. In the first step, the proof closely follows the proof outline of [10] in which we construct an upper bound on the relative entropy $D(P^{(\mathscr{C}_n)}_{\mathbf{V}} \| Q^n_V)$ based on a typical set construction of n-symbol sequences. In the second step, the proof diverges from [10] to analyze how the relative entropy upper bound concentrates. This second step uses the k-wise independent property of the random code $\mathscr{C}_n$.

Define the information density of a scalar pair $(u, v) \in \mathcal{U} \times \mathcal{V}$ as $i_{Q_{U,V}}(u; v) \triangleq \log_2 \frac{Q_{V|U}(v|u)}{Q_V(v)}$. In turn, define the information density of an n-symbol sequence pair $(\mathbf{u}, \mathbf{v}) \in \mathcal{U}^n \times \mathcal{V}^n$,

$$i_{Q^n_{U,V}}(\mathbf{u}; \mathbf{v}) \triangleq \sum_{j=1}^{n} i_{Q_{U,V}}(u_j; v_j).$$

For $\epsilon > 0$, define a typical set of n-symbol sequence pairs

$$\mathcal{A}_\epsilon \triangleq \left\{ (\mathbf{u}, \mathbf{v}) \in \mathcal{U}^n \times \mathcal{V}^n : i_{Q^n_{U,V}}(\mathbf{u}; \mathbf{v}) < (I(U; V) + \epsilon) n \right\}.$$

Recall that for an $[n, R'n]$ code $\mathcal{C}_n$, the PMF $P^{(\mathcal{C}_n)}_{\mathbf{V}}$ is the PMF of $\mathbf{V}$ when $\mathbf{U}$ is a codeword drawn from the code $\mathcal{C}_n$ (c.f. (8)). We split $P^{(\mathcal{C}_n)}_{\mathbf{V}}$ into two terms based on the typical set $\mathcal{A}_\epsilon$: for $\mathbf{v} \in \mathcal{V}^n$, define

$$P^{(\mathcal{C}_n)}_{\mathbf{V},1}(\mathbf{v}) \triangleq$$
$$2^{-Rn} \sum_{w \in \mathcal{W}} Q^n_{V|U}(\mathbf{v}|\mathbf{u}(w, \mathcal{C}_n)) \mathbb{1}\{(\mathbf{u}(w, \mathcal{C}_n), \mathbf{v}) \in \mathcal{A}_\epsilon\},$$

and define

$$P^{(\mathcal{C}_n)}_{\mathbf{V},2}(\mathbf{v}) \triangleq$$
$$2^{-Rn} \sum_{w \in \mathcal{W}} Q^n_{V|U}(\mathbf{v}|\mathbf{u}(w, \mathcal{C}_n)) \mathbb{1}\{(\mathbf{u}(w, \mathcal{C}_n), \mathbf{v}) \notin \mathcal{A}_\epsilon\}.$$

By inspection, $P^{(\mathcal{C}_n)}_{\mathbf{V}} = P^{(\mathcal{C}_n)}_{\mathbf{V},1} + P^{(\mathcal{C}_n)}_{\mathbf{V},2}$; note that $P^{(\mathcal{C}_n)}_{\mathbf{V},1}$ and $P^{(\mathcal{C}_n)}_{\mathbf{V},2}$ may not be PMFs. We also define the ratios

$$\Delta^{(\mathcal{C}_n)}_{\mathbf{V},1}(\mathbf{v}) \triangleq \frac{P^{(\mathcal{C}_n)}_{\mathbf{V},1}(\mathbf{v})}{Q^n_V(\mathbf{v})} \quad \text{and} \quad \Delta^{(\mathcal{C}_n)}_{\mathbf{V},2}(\mathbf{v}) \triangleq \frac{P^{(\mathcal{C}_n)}_{\mathbf{V},2}(\mathbf{v})}{Q^n_V(\mathbf{v})}.$$

We restate a result from [10] that bounds the relative entropy of $P^{(\mathcal{C}_n)}_{\mathbf{V}}$ and Q^n_V in terms of the introduced quantities.

Lemma 4 ([10, Lemma 3]). *For every $[n, R'n]$ code $\mathcal{C}_n$,*

$$D\left(P^{(\mathcal{C}_n)}_{\mathbf{V}} \middle\| Q^n_V \right) \leq H_2\left(\sum_{\mathbf{v} \in \mathcal{V}^n} P^{(\mathcal{C}_n)}_{\mathbf{V},2}(\mathbf{v}) \right)$$
$$+ D\left(P^{(\mathcal{C}_n)}_{\mathbf{V},1} \middle\| Q^n_V \right) + D\left(P^{(\mathcal{C}_n)}_{\mathbf{V},2} \middle\| Q^n_V \right).$$

We remark that the RHS of the inequality of Lemma 4 is well defined if we extend the definition of relative entropy $D(\cdot\|\cdot)$ in the natural way to account for functions $P^{(\mathcal{C}_n)}_{\mathbf{V},1}$ and $P^{(\mathcal{C}_n)}_{\mathbf{V},2}$ which may not be PMFs. The following sufficient condition for Lemma 3 follows from Lemma 4.

Lemma 5 (Sufficient Condition for Lemma 3). *Suppose that for some $\pi_0 \in [0, 1]$ and with probability at least $1 - \pi_0$ over the random code distribution, for some $\pi_1 > 0$*

$$\sum_{\mathbf{v} \in \mathcal{V}^n} P^{(\mathscr{C}_n)}_{\mathbf{V},2}(\mathbf{v}) < 2^{-\pi_1 n} \tag{9}$$

and

$$\Delta^{(\mathscr{C}_n)}_{\mathbf{V},1}(\mathbf{v}) < 1 + 2^{-\pi_1 n} \text{ for all } \mathbf{v} \in \mathcal{V}^n. \tag{10}$$

Then

$$\mathbb{P}_{\mathscr{C}_n}\left(D\left(P^{(\mathscr{C}_n)}_{\mathbf{V}} \middle\| Q^n_V \right) \geq q_n 2^{-\pi_1 n} \right) \leq \pi_0 \tag{11}$$

where $q_n = 2\log_2 e + \pi_1 n + n\log_2\left(\max_{v\in\mathrm{supp}(Q_V)}\frac{1}{Q_V(v)}\right).$

Proof of Lemma 5. Let $\pi_1 > 0$ and suppose that $\mathcal{C}_n$ is a realization of $\mathscr{C}_n$ such that both (9) and (10) hold. We bound each of the 3 terms in the inequality of Lemma 4 using (9) and (10).

Consider the first term. Following (9) and the inequality[7] $H_2(x) \le x\log_2\frac{e}{x}$ for $x\in[0,1]$, we have that

$$H_2\left(\sum_{v\in\mathcal{V}^n} P_{V,2}^{(\mathcal{C}_n)}(v)\right) \le H_2(2^{-\pi_1 n}) < 2^{-\pi_1 n}(\log_2 e + \pi_1 n).$$
(12)

Moving on to the second term, following (10) and the inequality $\log_2(1+x) \le x\log_2 e$ for $x > 0$, we have that

$$\begin{aligned} D(P_{V,1}^{(\mathcal{C}_n)}\|Q_V^n) &\triangleq \sum_{v\in\mathcal{V}^n} P_{V,1}^{(\mathcal{C}_n)}(v)\log_2\Delta_{V,1}^{(\mathcal{C}_n)} \\ &< \sum_{v\in\mathcal{V}^n} P_{V,1}^{(\mathcal{C}_n)}\log_2(1 + 2^{-\pi_1 n}) \\ &\le \log_2(1 + 2^{-\pi_1 n}) \le 2^{-\pi_1 n}\log_2 e. \end{aligned}$$
(13)

Moving to the last term, we will use the following inequality which uses the assumption that $|\mathcal{V}|$ is finite: $\Delta_{V,2}^{(\mathcal{C}_n)}(v) \triangleq \frac{P_{V,2}^{(\mathcal{C}_n)}(v)}{Q_V^n(v)} \le \max_{v'\in\mathrm{supp}(Q_V^n)}\frac{1}{Q^n(v')} = (\max_{v'\in\mathrm{supp}(Q_V)}\frac{1}{Q(v')})^n$ for all $v\in\mathcal{V}^n$. Following this inequality and (9), we have that

$$\begin{aligned} D(P_{V,2}^{(\mathcal{C}_n)}\|Q_V^n) &\triangleq \sum_{v\in\mathcal{V}^n} P_{V,2}^{(\mathcal{C}_n)}(v)\log_2\Delta_{V,2}^{(\mathcal{C}_n)} \\ &\le \sum_{v\in\mathcal{V}^n} P_{V,2}^{(\mathcal{C}_n)}(v)n\log_2\left(\max_{v'\in\mathrm{supp}(Q_V)}\frac{1}{Q_V(v')}\right) \\ &< 2^{-\pi_1 n}n\log_2\left(\max_{v'\in\mathrm{supp}(Q_V)}\frac{1}{Q_V(v')}\right). \end{aligned}$$
(14)

Combining the bounds (12), (13) and (14) together with Lemma Lemma 4, the desired inequality (11) immediately follows. ∎

In the remainder of the proof of Lemma 3, we apply the framework of the sufficient condition (Lemma 10) and show that inequalities (9) and (10) hold with probability $1 - \pi_0$ over the distribution of $\mathscr{C}_n$ for a value $\pi_0 = 2^{-k\Omega(n)+n\log_2|\mathcal{V}|}$ and some $\pi_1 > 0$. As the primary technical tools of the proof, we use the concentration inequalities of Schmidt, Siegel and Srinivasan [25] and Bellare and Rompel [26] for sums of k-wise independent random variables.

B. Proof of Lemma 3

First, we show that inequality (9) holds with high probability over the random code $\mathscr{C}_n$ for some $\pi_1 > 0$. Consider the

[7]This inequality follows from an application of both the inequality $\frac{x}{1+x} \le \ln(1+x)$ for $x > -1$ and the definition of $H_2(x)$.

quantity

$$\begin{aligned} &\sum_{v\in\mathcal{V}^n} P_{V,2}^{(\mathscr{C}_n)}(v) \\ &= \sum_{w\in\mathcal{W}} 2^{-R'n}\sum_{v\in\mathcal{V}^n} Q_{V|U}^n(v|U(w,\mathscr{C}_n))\mathbb{1}\left\{(U(w,\mathscr{C}_n), v)\notin\mathcal{A}_\epsilon\right\} \\ &= \sum_{w\in\mathcal{W}} 2^{-R'n}\mathbb{P}_{V\sim Q_{V|U}^n}\left((U(w,\mathscr{C}_n),V)\notin\mathcal{A}_\epsilon\Big|U = U(w,\mathscr{C}_n)\right) \end{aligned}$$
(15)

Note that (15) is a sum of $|\mathcal{W}| = 2^{R'n}$ k-wise-independent terms following that the codewords of $\mathscr{C}_n$ are k-wise independent.

For $w\in\mathcal{W}$, the expectation of the wth term in the sum of (15) is

$$\begin{aligned} &2^{-R'n}\mathbb{E}_{\mathscr{C}_n}\mathbb{P}_{V\sim Q_{V|U}^n}\left((U(w,\mathscr{C}_n),V)\notin\mathcal{A}_\epsilon\Big|U = U(w,\mathscr{C}_n)\right) \\ &\overset{(a)}{=} 2^{-R'n}\mathbb{P}_{(U,V)\sim Q_{U,V}^n}\left((U,V)\notin\mathcal{A}_\epsilon\right) \\ &\overset{(b)}{=} 2^{-R'n}\mathbb{P}_{(U,V)\sim Q_{U,V}^n}\left(i_{Q_{U,V}^n}(U;V) \ge (I(U;V)+\epsilon)n\right) \\ &\overset{(c)}{=} 2^{-R'n}\mathbb{P}_{(U,V)\sim Q_{U,V}^n}\left(2^{\lambda i_{Q_{U,V}^n}(U;V)} \ge 2^{\lambda(I(U;V)+\epsilon)n}\right) \\ &\overset{(d)}{\le} 2^{-R'n}\left(\frac{\mathbb{E}_{(U,V)\sim Q_{U,V}}\left[2^{\lambda i_{Q_{U,V}}(U;V)}\right]}{2^{\lambda(I(U;V)+\epsilon)}}\right)^n \\ &= 2^{-\lambda\left(I(U;V)+\epsilon-\frac{1}{\lambda}\log_2\mathbb{E}_{(U,V)\sim Q_{U,V}}\left[2^{\lambda i_{Q_{U,V}}(U;V)}\right]\right)n-R'n} \\ &\overset{(e)}{=} 2^{-\lambda(I(U;V)+\epsilon-D_{\lambda+1}(Q_{U,V}\|Q_U Q_V))n-R'n} \\ &= 2^{-(\alpha_{\lambda,\epsilon}+R')n} \end{aligned}$$
(16)

where (a) follows from the fact that $U(w,\mathscr{C}_n)$ is distributed as Q_U^n, (b) follows from the definition of $\mathcal{A}_\epsilon$, (c) holds for any $\lambda > 0$, (d) follows from Markov's inequality and the product form of the joint PMF $Q_{U,V}^n$, (e) follows from the definition of Rényi divergence of order $\lambda+1$, and where $\alpha_{\lambda,\epsilon} \triangleq \lambda\left(I(U;V)+\epsilon-D_{\lambda+1}(Q_{U,V}\|Q_U Q_V)\right)$.

For $\epsilon > 0$, we remark that i) $\alpha_{\lambda,\epsilon}$ tends to 0 as λ tends to 0, and ii) $\alpha_{\lambda,\epsilon}$ is positive for small enough $\lambda > 0$; these follow from the facts that $D_{\lambda+1}(Q_{U,V}\|Q_U Q_V)$ is a continuous and non-decreasing function of $\lambda > 0$ and that $D_1(Q_{U,V}\|Q_U Q_V) = I(U;V)$. In the sequel, for a given $\epsilon > 0$, we let $\lambda > 0$ be small enough such that $\alpha_{\lambda,\epsilon}\in(0,R')$. Moving forward, we write $\alpha_{\lambda,\epsilon}$ as simply α when the dependency on λ and ϵ is clear from context.

Lemma 6 ([25, Theorem 3]). *Suppose that $\{T_w\}_{w\in\mathcal{W}}$ are random variables that take values in $[0,1]$, and define $T \triangleq \sum_{w\in\mathcal{W}} T_w$ and $\mu \triangleq \mathbb{E}[T]$. For $\tau > 0$, if the variables are k-wise independent for some $k \ge k^*(|\mathcal{W}|,\mu,\tau) \triangleq \lceil\frac{\mu\tau}{1-\frac{\mu}{|\mathcal{W}|}}\rceil$, then*

$$\mathbb{P}\left(T \ge \mu(1+\tau)\right) \le \frac{\binom{|\mathcal{W}|}{k^*}\left(\frac{\mu}{|\mathcal{W}|}\right)^{k^*}}{\binom{\mu(1+\tau)}{k^*}}.$$

Using the framework of Lemma 6, we set T_w for each $w\in\mathcal{W}$ to be the wth term in the sum of (15), i.e.,

$$T_w = 2^{-R'n}\mathbb{P}_{V\sim Q_{V|U}^n}\left((U(w,\mathscr{C}_n),V)\notin\mathcal{A}_\epsilon|U = U(w,\mathscr{C}_n)\right),$$

and in turn, we have that $T \triangleq \sum_{w \in T_w} T_w = \sum_{\boldsymbol{v} \in \mathcal{V}^n} P_{\boldsymbol{V},1}^{(\mathscr{C}_n)}(\boldsymbol{v})$. Note that the expectation $\mu \triangleq \mathbb{E}_{\mathscr{C}_n}[T]$ is bounded above by $2^{-\alpha n}$ following (16). For a parameter $\beta \in (0, \alpha)$ that will be set later, set τ such that $\mu(1+\tau) = 2^{(\beta-\alpha)n}$.

Before applying Lemma 6, we normalize the random variables $\{T_w\}_{w \in \mathcal{W}}$ to optimize the parameter k^*. For some parameter $\theta \in (0,1]$ which we will soon set, define $T'_w = \theta 2^{R'n} T_w$ and note that $T'_w \in [0,1]$. Similarly, define the normalized sum $T' = \theta 2^{R'n} T$, its normalized expectation $\mu' = \theta 2^{R'n} \mu$ which is bounded above by $\theta 2^{(R'-\alpha)n}$, and note that $\mu'(1+\tau) = \theta 2^{(R'+\beta-\alpha)}$. Now consider the quantity $k^*(|\mathcal{W}|, \mu', \tau)$ as a function of θ, and let n be large enough and choose $\theta \in (0,1]$ such that $k^*(|\mathcal{W}|, \mu', \tau)$ is equal to k; such a choice exists for fixed k and large enough n since $k^*(|\mathcal{W}|, \mu', \tau) \geq \mu'\tau = \theta 2^{(R'+\beta-\alpha)n} - \mu' \geq \theta 2^{(R'-\alpha)n}(2^{\beta n} - 1)$ is tending larger than k for fixed $\theta > 0$ as n tends to infinity following $\alpha < R'$.

We apply Lemma 6 to the normalized random variables $\{T'_w\}_{w \in \mathcal{W}}$. We have for large enough n

$$
\begin{aligned}
\mathbb{P}_{\mathscr{C}_n}\left(\sum_{\boldsymbol{v} \in \mathcal{V}^n} P_{\boldsymbol{V},2}^{(\mathscr{C}_n)}(\boldsymbol{v}) \geq 2^{(\beta-\alpha)n}\right) &= \mathbb{P}_{\mathscr{C}_n}\left(T \geq 2^{(\beta-\alpha)n}\right) \\
&\overset{(f)}{=} \mathbb{P}_{\mathscr{C}_n}\left(T' \geq \theta 2^{(R+\beta-\alpha)n}\right) \\
&\overset{(g)}{\leq} \frac{\binom{2^{R'n}}{k}\left(\frac{\mu'}{2^{R'n}}\right)^k}{\binom{\theta 2^{(R'+\beta-\alpha)n}}{k}} \\
&\overset{(h)}{\leq} \frac{k^k}{k!}\left(\frac{\mu'}{\theta 2^{(R'+\beta-\alpha)n}}\right)^k \\
&\overset{(i)}{\leq} \frac{k^k}{k!} 2^{-k\beta n}
\end{aligned}
\tag{17}
$$

where (f) follows from the normalization $T' = \theta 2^{R'n} T$, (g) follows for large enough n from Lemma 6 and the choice of θ such that $k^* = k$, (h) follows from the inequalities $\frac{m^k}{k^k} \leq \binom{m}{k} \leq \frac{m^k}{k!}$ for any $1 \leq k \geq m$, and (i) follows from the bound $\mu' \leq \theta 2^{(R'-\alpha)}$.

Next, we show that inequality (10) holds with high probability over the random code $\mathscr{C}_n$. For $\boldsymbol{v} \in \mathcal{V}^n$, expand $\Delta_{\boldsymbol{V},1}^{(\mathscr{C}_n)}(\boldsymbol{v})$:

$$
\begin{aligned}
\Delta_{\boldsymbol{V},1}^{(\mathscr{C}_n)}(\boldsymbol{v}) &\triangleq \frac{P_{\boldsymbol{V},1}^{(\mathscr{C}_n)}(\boldsymbol{v})}{Q_V^n(\boldsymbol{v})} \\
&= \sum_{w \in \mathcal{W}} 2^{-R'n} \frac{Q_{V|U}^n(\boldsymbol{v}|\boldsymbol{U}(w,\mathscr{C}_n))}{Q_V^n(\boldsymbol{v})} \mathbb{1}\{(\boldsymbol{U}(w,\mathscr{C}_n), \boldsymbol{v}) \in \mathcal{A}_\epsilon\}.
\end{aligned}
\tag{18}
$$

Note that (18) is a sum of $|\mathcal{W}| = 2^{R'n}$ k-wise independent terms following that the codewords of $\mathscr{C}_n$ are k-wise independent. For $w \in \mathcal{W}$, the expectation of the wth term in the sum

of (18) is

$$
\begin{aligned}
&2^{-R'n} \mathbb{E}_{\mathscr{C}_n}\left[\frac{Q_{V|U}^n(\boldsymbol{v}|\boldsymbol{U}(w,\mathscr{C}_n))}{Q_V^n(\boldsymbol{v})} \mathbb{1}\{(\boldsymbol{U}(w,\mathscr{C}_n), \boldsymbol{v}) \in \mathcal{A}_\epsilon\}\right] \\
&\overset{(j)}{\leq} 2^{-R'n} \mathbb{E}_{\mathscr{C}_n}\left[\frac{Q_{V|U}^n(\boldsymbol{v}|\boldsymbol{U}(w,\mathscr{C}_n))}{Q_V^n(\boldsymbol{v})}\right] \\
&\overset{(k)}{=} 2^{-R'n} \sum_{\boldsymbol{u} \in \mathcal{U}^n} Q_U^n(\boldsymbol{u}) \frac{Q_{V|U}^n(\boldsymbol{v}|\boldsymbol{u})}{Q_V^n(\boldsymbol{v})} \\
&= 2^{-R'n}
\end{aligned}
\tag{19}
$$

where (j) follows from the trivial bound $\mathbb{1}\{\cdot\} \leq 1$ and (k) follows from the distribution of codeword $\boldsymbol{U}(w,\mathscr{C}_n) \sim Q_U^n$.

Lemma 7 ([26, Lemma 2.3]). *Let $k \geq 4$ be an even integer. Suppose that $\{T_w\}_{w \in \mathcal{W}}$ are k-wise independent random variables that take values in $[0,1]$, and define $T \triangleq \sum_{w \in \mathcal{W}} T_w$ and $\mu \triangleq \mathbb{E}[T]$. For any $\tau > 0$,*

$$
\mathbb{P}(T \geq \mu(1+\tau)) \leq 8\left(\frac{k\mu + k^2}{(\mu\tau)^2}\right)^{k/2}.
$$

Using the framework of Lemma 7, fix $\boldsymbol{v} \in \mathcal{V}^n$ and set T_w for each $w \in \mathcal{W}$ to be

$$
T_w = 2^{(-I(U;V)-\epsilon)n}\left(\frac{Q_{V|U}^n(\boldsymbol{v}|\boldsymbol{U}(w,\mathscr{C}_n))}{Q_V^n(\boldsymbol{v})}\right) \mathbb{1}\{(\boldsymbol{U}(w,\mathscr{C}_n), \boldsymbol{v}) \in \mathcal{A}_\epsilon\}
$$

which coincides with the wth term in the sum of (18) normalized by the factor $2^{(R'-I(U;V)-\epsilon)n}$. This normalization factor was chosen to ensure T_w is bounded above by 1 which follows from that fact that for any $(\boldsymbol{u}, \boldsymbol{v}) \in \mathcal{A}_\epsilon$ we have that $\frac{Q_{V|U}^n(\boldsymbol{v}|\boldsymbol{u})}{Q_V^n(\boldsymbol{v})} < 2^{(I(U;V)+\epsilon)n}$. Set $T = \sum_{w \in \mathcal{W}} T_w$ and note that $\mu \triangleq \mathbb{E}_{\mathscr{C}_n}[T]$ is bounded above by $2^{(R'-I(U;V)-\epsilon)n}$ following (19) and the choice of normalization factor. Finally, set τ such that $\mu(\tau + 1) = 2^{(R'-I(U;V)-\epsilon)n}(1 + 2^{(\beta-\alpha)n})$ and note that $\mu\tau = 2^{(R'-I(U;V)-\epsilon)n}(1 + 2^{(\beta-\alpha)n}) - \mu \geq 2^{(R'-I(U;V)-\epsilon+\beta-\alpha)n}$. Applying Lemma 7, we have that for for even integer $k \geq 4$, small enough $\epsilon > 0$ and large enough n

$$
\begin{aligned}
\mathbb{P}_{\mathscr{C}_n}\left(\Delta_{\boldsymbol{V},1}^{(\mathscr{C}_n)}(\boldsymbol{v}) \geq 1 + 2^{(\beta-\alpha)n}\right) &= \mathbb{P}_{\mathscr{C}_n}\left(T \geq \mu(1+\tau)\right) \\
&\overset{(\ell)}{\leq} 8\left(\frac{k2^{(R'-I(U;V)-\epsilon)n} + k^2}{2^{2(R'-I(U;V)-\epsilon+\beta-\alpha)n}}\right)^{k/2} \\
&\overset{(m)}{\leq} 8\left(\frac{(k+1)2^{(R'-I(U;V)-\epsilon)n}}{2^{2(R'-I(U;V)-\epsilon+\beta-\alpha)n}}\right)^{k/2} \\
&= 8(k+1)^{k/2} \cdot 2^{-k\eta n}.
\end{aligned}
\tag{20}
$$

where (ℓ) follows from Lemma 7 and the bounds $\mu \leq 2^{(R'-I(U;V)-\epsilon)n}$ and $\mu\tau \geq 2^{(R'-I(U;V)-\epsilon+\beta-\alpha)n}$, and (m) follows for small enough $\epsilon > 0$ and large enough n such that $k2^{(R'-I(U;V)-\epsilon)n} >> k^2$, and where

$$
\eta = \frac{R' - I(U;V) - \epsilon + 2(\beta - \alpha)}{2}
\tag{21}
$$

In turn, by a simple union bound over all $\boldsymbol{v} \in \mathcal{V}^n$, and by letting $k \geq 4$ be an even integer, $\epsilon > 0$ be small enough and n be large enough,

$$\mathbb{P}_{\mathscr{C}_n}\left(\exists \boldsymbol{v} \in \mathcal{V}^n \text{ s.t. } \Delta_{\boldsymbol{V},1}^{(\mathscr{C}_n)}(\boldsymbol{v}) \geq 1 + 2^{(\beta-\alpha)n}\right)$$
$$\leq 8k(k+1)^{k/2} \cdot 2^{-(k\eta_1 + \log_2 |\mathcal{V}|)n}. \tag{22}$$

To complete the proof, we put together the above results and apply the sufficient condition (Lemma 3). In the framework of Lemma 3, we set $\pi_1 = \alpha - \beta$. If $\pi_1 > 0$, then it follows from Lemma 3 that the inequalities (9) and (10) hold with probability at least $1 - \pi_0$ where

$$\pi_0 = \frac{k^k}{k!} 2^{-k\beta n} + 8k(k+1)^{k/2} \cdot 2^{(-k\eta + \log_2 |\mathcal{V}|)n}$$

where the expression for π_0 follows from (17) and (22) together with a simple union bound.

The last step is to show that for some choice of the free parameters $\epsilon > 0$, $\lambda > 0$ and $\beta \in (0, \alpha)$ we have that $\pi_1 > 0$ and $\pi_0 = 2^{-k\Omega(n) + n \log_2 |\mathcal{V}|}$. Recall that for a fixed $\epsilon > 0$, $\alpha = \alpha_{\lambda,\epsilon}$ tends to 0 as λ tends to 0, and $\alpha_{\lambda,\epsilon}$ is positive for small enough $\lambda > 0$. Furthermore, recall that $R' > I(U;V)$ by assumption, and thus, η given by (21) is positive for small enough $\epsilon > 0$, small enough $\alpha_{\lambda,\epsilon} > 0$, and any $\beta \in (0, \alpha_{\lambda,\epsilon})$. Thus, given even $k \geq 4$, we can pick $\epsilon > 0$ small enough, and in turn, pick $\lambda > 0$ small enough such that both $\alpha_{\lambda,\epsilon}$ and η_1 are positive. In turn, picking $\beta \in (0, \alpha_{\lambda,\epsilon})$ ensures that $\alpha_{\lambda,\epsilon} - \beta > 0$ and thus $\pi_1 > 0$. Thus, $\pi_0 = 2^{-k\Omega(n) + \log_2 |\mathcal{V}|}$. This completes the proof of Lemma 5.

V. PROOF OF THEOREM 2

Setup: Let $p \in [0, 1/2]$ and $r \in [0, 1]$ such that $1 - H_2(p) - r > 0$. For $\epsilon > 0$ and $\epsilon' \in (0, \epsilon)$, let $R = 1 - H_2(p) - r - \epsilon$ and $R' = r + \epsilon'$. Let k be a positive integer to be set in the proof. The goal of the proof is to show that for large enough k constant in n and for large enough n, there exists an $[n, Rn, R'n, k]$ pseudolinear code $\mathcal{C}_n$ such that both $\text{Sem}(\mathcal{C}_n) = 2^{-\Omega(n)}$ and $P_{\text{error}}^{\max}(\mathcal{C}_n) = o(1)$.

Encoding: Alice uses an $[n, Rn, R'n]$ code $\mathcal{C}_n = \{\boldsymbol{x}(m,w)\}_{(m,w) \in \mathcal{M} \times \mathcal{W}}$ to encode her message M. That is, for a message distribution $P_M \in \mathcal{P}(\mathcal{M})$, Alice draws $M \sim P_M$ and $W \sim \text{Unif}(\mathcal{W})$ and transmits $\boldsymbol{x}(M,W)$.

Decoding: Upon receiving the channel output $\boldsymbol{y}$, Bob performs min-distance decoding by choosing the message estimate $\widehat{m}$ and key estimate $\widehat{w}$ such that

$$(\widehat{m}, \widehat{w}) = \arg \min_{(m,w) \in \mathcal{M} \times \mathcal{W}} d_H(\boldsymbol{x}(m,w), \boldsymbol{y})$$

where d_H denotes the Hamming distance.

A. Code Distribution

We show the existence of a good code via a random coding argument. As our random code distribution, we will use the following distribution over $[n, Rn, R'n, k]$ pseudolinear codes.

Definition 2 (Random Code Dist.). *Let $F[n, Rn, R'n, k]$ be the distribution over all $[n, Rn, R'n, k]$ pseudolinear codes*
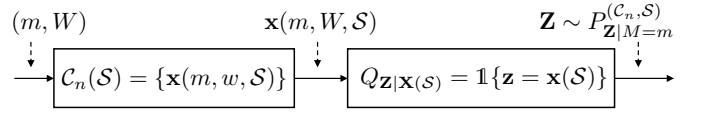

Fig. 3: A mapping of the quantities in (25) to the soft-covering problem of Fig. 2.

where the parity check matrix H (c.f. Definition 1) is fixed and the generator matrix G is chosen uniformly from $\{0,1\}^{\ell \times n}$.

The following property of $F[n, Rn, R'n, k]$ is useful.

Lemma 8 ([15, Lemma 9.1]). *The codewords of $\mathscr{C}_n \sim F[n, Rn, R'n, k]$ are uniformly distributed over $\{0,1\}^n$ and are k-wise independent.*

B. Secrecy Analysis

For a given $\mathcal{S} \in \mathscr{S}$, let $Q_{\boldsymbol{Z}}^{(\mathcal{S})}$ denote the PMF of the adversary's observation $\boldsymbol{Z} \in \{0,1\}^{rn}$ when Alice sends a *random n-bit sequence* $\boldsymbol{X} \sim Q_X^n \triangleq \text{Unif}(\{0,1\}^n)$ through the channel. We have that

$$Q_{\boldsymbol{Z}}^{(\mathcal{S})}(\boldsymbol{z}) = Q_{\boldsymbol{X}(\mathcal{S})}(\boldsymbol{z}) = Q_X^{rn}(\boldsymbol{z}), \text{ for all } \boldsymbol{z} \in \{0,1\}^{rn}. \tag{23}$$

Furthermore, for an $[n, Rn, R'n]$ code $\mathcal{C}_n$, let $P_{M,\boldsymbol{Z}}^{(\mathcal{C}_n,\mathcal{S})}$ denote the joint PMF of message M and observation $\boldsymbol{Z}$ when Alice sends the codeword $\boldsymbol{x}(M,W,\mathcal{C}_n)$ through the channel. Then for marginal PMF $P_M \in \mathcal{P}(\mathcal{M})$,

$$I_{\mathcal{C}_n}(M;\boldsymbol{Z}) \triangleq D\left(P_{M,\boldsymbol{Z}}^{(\mathcal{C}_n,\mathcal{S})} \| P_M P_{\boldsymbol{Z}}^{(\mathcal{C}_n,\mathcal{S})}\right)$$
$$\overset{(a)}{=} D\left(P_{M,\boldsymbol{Z}}^{(\mathcal{C}_n,\mathcal{S})} \| P_M Q_{\boldsymbol{Z}}^{(\mathcal{S})}\right) - D\left(P_{\boldsymbol{Z}}^{(\mathcal{C}_n,\mathcal{S})} \| Q_{\boldsymbol{Z}}^{(\mathcal{S})}\right)$$
$$\overset{(b)}{\leq} D\left(P_{M,\boldsymbol{Z}}^{(\mathcal{C}_n,\mathcal{S})} \| P_M Q_{\boldsymbol{Z}}^{(\mathcal{S})}\right)$$
$$\leq \sum_{m \in \mathcal{M}} P_M(m) \max_{m' \in \mathcal{M}} D\left(P_{\boldsymbol{Z}|M=m'}^{(\mathcal{C}_n,\mathcal{S})} \| Q_{\boldsymbol{Z}}^{(\mathcal{S})}\right)$$
$$= \max_{m \in \mathcal{M}} D\left(P_{\boldsymbol{Z}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \| Q_{\boldsymbol{Z}}^{(\mathcal{S})}\right) \tag{24}$$

where (a) follows from the relative entropy chain rule and (b) follows from the property $D(\cdot\|\cdot) \geq 0$. Thus,

$$\text{Sem}(\mathcal{C}_n) = \max_{P_M \in \mathcal{P}(\mathcal{M}), \mathcal{S} \in \mathscr{S}} I_{\mathcal{C}_n}(M;\boldsymbol{Z})$$
$$\overset{(c)}{\leq} \max_{\mathcal{S} \in \mathscr{S}} \max_{m \in \mathcal{M}} D\left(P_{\boldsymbol{Z}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \| Q_{\boldsymbol{Z}}^{(\mathcal{S})}\right)$$
$$\overset{(d)}{=} \max_{\mathcal{S} \in \mathscr{S}} \max_{m \in \mathcal{M}} D\left(P_{\boldsymbol{Z}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \| Q_X^{rn}\right) \tag{25}$$

where (c) follows from (24) and (d) follows from (23).

Consider the relative entropy $D\left(P_{\boldsymbol{Z}|M=m}^{(\mathcal{C}_n,\mathcal{S})} \| Q_X^{rn}\right)$ in the framework of the soft-covering lemma for k-wise independent codewords (Lemma 3), as illustrated in Fig. 3. Here, (m,W) is uniformly drawn from a message-key product set $\{m\} \times \mathcal{W}$ of rate R'/r, i.e., $|\{m\} \times \mathcal{W}| = 2^{R'n} = 2^{rn\frac{R'}{r}}$. Since rate $R'/r = (r+\epsilon')/r$ is greater than $I(X;Z) = 1$, it follows from Lemma 3 that there exists $\gamma_0 > 0$ and $\gamma_1 > 0$ such that for even integer $k \geq 4$ and large enough n,

$$\mathbb{P}_{\mathscr{C}_n}\left(D\left(P_{\boldsymbol{Z}|M=m}^{(\mathscr{C}_n,\mathcal{S})} \| Q_X^{rn}\right) > 2^{-\gamma_1 rn}\right) \leq 2^{(-k\gamma_0 + 1)rn}. \tag{26}$$

In turn,

$$\mathbb{P}_{\mathscr{C}_n}\left(\mathrm{Sem}(\mathscr{C}_n) > 2^{-\gamma_1 rn}\right)$$

$$\overset{(e)}{\leq} \mathbb{P}_{\mathscr{C}_n}\left(\max_{\mathcal{S} \in \mathscr{S}} \max_{m \in \mathcal{M}} D\left(P_{\boldsymbol{Z}|M=m}^{(\mathscr{C}_n,\mathcal{S})} || Q_X^{rn}\right) > 2^{-\gamma_1 rn}\right)$$

$$\leq \mathbb{P}_{\mathscr{C}_n}\left(\bigcup_{\mathcal{S} \in \mathscr{S}} \bigcup_{m \in \mathcal{M}} \left\{D\left(P_{\boldsymbol{Z}|M=m}^{(\mathscr{C}_n,\mathcal{S})} || Q_X^{rn}\right) > 2^{-\gamma_1 rn}\right\}\right)$$

$$\overset{(f)}{\leq} 2^{(-k\gamma_0 r + r + R + 1)n} \tag{27}$$

where (e) follows from (25), and (f) follows for large enough n from a simple union bound, the inequality $|\mathscr{S}| = \binom{n}{rn} \leq 2^n$ and (26).

C. Reliability Analysis

Unlike the above secrecy analysis, the reliability analysis requires additional structure of the code $\mathscr{C}_n$ beyond the k-wise independence property. In particular, we will use the pseudo-linear structure of $\mathscr{C}_n$. We restate a reliability result of [21] without proof. For a code $\mathcal{C}_n$ and a message $m \in \mathcal{M}$, define the probability of decoding error conditioned on $M = m$ as

$$P_{\mathrm{error}}^{(m)}(\mathcal{C}_n) \triangleq \mathbb{P}(\widehat{M} \neq m | M = m)$$

where the probability is w.r.t. $W \sim \mathrm{Unif}(\mathcal{W})$ and the adversary's choice of bit read/flip locations.

Lemma 9 ([21, Theorem 1]). *Suppose that $p \in (0, 1/2)$ and $r < 1 - H_2(p)$. If the key rate $R' > r$ and the sum rate $R + R' < 1 - H_2(p)$, then for large enough (but fixed) k and any fixed $\delta > 0$, there exists $\gamma_2 > 0$ such that for large enough n and any $m \in \mathcal{M}$,*

$$\mathbb{P}_{\mathscr{C}_n}\left(P_{\mathrm{error}}^{(m)}(\mathscr{C}_n) > \delta\right) \leq 2^{-k\gamma_2 n}. \tag{28}$$

We apply Lemma 9 to bound the maximum probability of error $P_{\mathrm{error}}^{\max}(\mathscr{C}_n) \triangleq \max_{m \in \mathcal{M}} P_{\mathrm{error}}^{(m)}(\mathscr{C}_n)$. Note that our choice of ϵ and ϵ' ensures that $R' > r$ and $R + R' < 1 - H_2(p)$. Also, we have that $R < 1 - H_2(p) - r$. Thus, for $\delta > 0$,

$$\mathbb{P}_{\mathscr{C}_n}\left(P_{\mathrm{error}(\mathscr{C}_n)}^{\max} > \delta\right) \triangleq \mathbb{P}_{\mathscr{C}_n}\left(\max_{m \in \mathcal{M}} P_{\mathrm{error}}^{(m)}(\mathscr{C}_n) > \delta\right)$$

$$\overset{(g)}{\leq} \sum_{m \in \mathcal{M}} \mathbb{P}_{\mathscr{C}_n}\left(P_{\mathrm{error}}^{(m)}(\mathscr{C}_n) > \delta\right)$$

$$\overset{(h)}{\leq} 2^{(-k\gamma_2 + 1 - H_2(p) - r)n} \tag{29}$$

where (g) follows from a union bound and (h) follows for large enough k and for large enough n via Lemma 9.

D. Combining Secrecy and Reliability Analysis

To complete the proof, we combine the secrecy and reliability analysis. For large enough k and k even, and for large enough n,

$$\mathbb{P}_{\mathscr{C}_n}\left(\{\mathrm{Sem}(\mathscr{C}_n) > 2^{-\gamma rn}\} \cup \{P_{\mathrm{error}}^{\max}(\mathscr{C}_n) > \delta\}\right)$$

$$\leq 2^{(-k\gamma_0 r + 2r + R)n} + 2^{(-k\gamma_2 + 1 - H_2(p) - r)n} \tag{30}$$

following both (27), (29) and a simple union bound. In summary, for large enough k and k even (which is constant in

n) and large enough n, we have that (30) is less than 1, and in turn, there exists an $[n, Rn, R'n, k]$ pseudolinear code $\mathcal{C}_n$ such that $\mathrm{Sem}(\mathcal{C}_n) \leq 2^{-\gamma_1 rn}$ and $P_{\mathrm{error}}^{\max}(\mathcal{C}_n) \leq \delta$.

VI. CONCLUSION

We showed that random pseudolinear codes achieve the best known lower bound of the semantic secrecy capacity of the binary adversarial wiretap channel of type II. A necessary condition on the non-linearity of a capacity achieving code was also shown. One possible avenue for future research is to apply further derandomization techniques to our random codes, e.g., in the spirit of [27]. The goal here is to replace random pseudolinear codes with a significantly derandomized class that can maintain the same error-correction and secrecy power while being more amendable to efficient decoding algorithms.

APPENDIX A
LINEAR COSET CODING SCHEMES

In this appendix, we prove that the linear coset coding scheme of Ozarow and Wyner [1] is not semantically-secret for any positive message rate. We first define coset coding.

The linear coset coding scheme, proposed in [1], is as follows: Let $R > 0$ be the message rate. For blocklength n, let H be the $Rn \times n$ parity check matrix of some $[n, n - Rn]$ binary linear code. *Encoding:* Suppose that Alice wants to transmit a message $m \in \{0, 1\}^{Rn}$. Alice encodes m by choosing the n bit codeword $\boldsymbol{x}$ randomly and uniformly from the set of solutions $\{\boldsymbol{x}' \in \{0, 1\}^n : \boldsymbol{x}' H^T = \boldsymbol{m}\}$ and transmits $\boldsymbol{x}$ over the (noiseless) $(0, r)$-AWTC II. *Decoding:* Upon receiving $\boldsymbol{x}$, Bob performs decoding by choosing the message estimate $\widehat{m} = \boldsymbol{x} H^T$. It is easy to show that the above linear coset coding scheme is an $[n, Rn, (1-R)n]$ linear code.

We prove the following result.

Lemma 10. *Let rate $R > 0$. For large enough n, any $[n, Rn, (1-R)n]$ binary code $\mathcal{C}_n$ that is a linear coset coding scheme has semantic leakage $\mathrm{Sem}(\mathcal{C}_n) \geq 1$.*

For any $R > 0$, let $\mathcal{C}_n$ be an $[n, Rn, (1-R)n]$ binary code that is a linear coset coding scheme and let H be the corresponding $Rn \times n$ parity check matrix. Suppose that Alice's message is uniformly distributed over $\{0, 1\}^n$. To prove Lemma 10, we will use the following result due to Ozarow and Wyner.

Lemma 11 ([1, Lemma 4]). *For an index set $\mathcal{I} \subseteq [n]$, let $H(\mathcal{I})$ denote the $|\mathcal{I}|$ columns of H indexed by $\mathcal{I}$. The adversary's equivocation is*

$$\Delta \triangleq \min_{\mathcal{S} \in \mathscr{S}} H(M|\boldsymbol{Z}) = \min_{\mathcal{I} \subseteq [n]: |\mathcal{I}| = (1-r)n} \mathrm{rank}\left(H(\mathcal{I})\right). \tag{31}$$

Recall the following definitional inequalities:

$$\mathrm{Sem}(\mathcal{C}_n) \geq \max_{\mathcal{S} \in \mathscr{S}} I_{\mathcal{S}}(M; \boldsymbol{Z})$$

$$= H(M) - \min_{\mathcal{S} \in \mathscr{S}} H(M|\boldsymbol{Z}) = Rn - \Delta.$$

Thus, to show that $\mathrm{Sem}(\mathcal{C}_n) \geq 1$ for large enough n, it is sufficient to show that $\Delta \leq Rn - 1$.

Let n be large enough and suppose by contradiction that $\Delta = Rn$. By Lemma 11, we have that $\mathrm{rank}(H(\mathcal{I})) = Rn$ for every set $\mathcal{I} \subseteq [n]$ s.t. $|\mathcal{I}| = (1-r)n$. This in turn by the definition of H implies that the $[n, (1-R)n]$ binary code with parity check matrix H has minimum distance, denoted $d_{\min}$, of at least $Rn + 1$. However, by the Plotkin bound of Lemma 2, we have that $1 - R \leq 1 - 2\frac{d_{\min}}{n} + o(1)$, or equivalently, $d_{\min} \leq \frac{Rn}{2} + o(n)$. Thus, for n large enough such that the $o(n)$ term is negligible, we have a contradiction. This completes the proof of Lemma 10.

APPENDIX B
DISCUSSION OF ASSUMPTION 1

We show that if the generator matrix G of an $[n, Rn]$ linear code $\mathcal{C}_n$ is not full rank, then either the probability of decoding error is large such that $P_{\mathrm{error}}^{\max}(\mathcal{C}_n) \geq 1/2$ or both $\mathcal{W}$ and G can be replaced with a smaller key set $\mathcal{W}'$ and generator matrix G', respectively, without changing the code. Let $\mathcal{C}_n$ be an $[n, Rn]$ linear code and suppose that G is not full rank.

Suppose that G_W is full rank. Since the channel is noiseless, Bob's received sequence is guaranteed to be a codeword in $\mathcal{C}_n$. Suppose that Bob receives the codeword $\boldsymbol{c} \in \mathcal{C}_n$. From Bob's perspective, the set of all possible message-key pairs that Alice could have sent is

$$\mathcal{M}_{\boldsymbol{c}} = \{(m, w) \in \mathcal{M} \times \mathcal{W} : \begin{bmatrix} m & w \end{bmatrix} G = \boldsymbol{c}\}$$
$$= \{(m, w) \in \mathcal{M} \times \mathcal{W} : mG_M + wG_W = \boldsymbol{c}\}.$$

Since the mapping $G : \{0,1\}^{(R+R')n} \rightarrow \{0,1\}^n$ is a linear transformation, the number of pairs in $\mathcal{M}_{\boldsymbol{c}}$ is $|\mathcal{M}_{\boldsymbol{c}}| = 2^{\mathrm{nullity}(G)} = 2^{(R+R')n - \mathrm{rank}(G)}$ where the second equality follows from the rank-nullity theorem. In turn, since $\mathrm{rank}(G) < (R+R')n$, it follows that $|\mathcal{M}_{\boldsymbol{c}}| \geq 2$. Now consider two unique pairs in $\mathcal{M}_{\boldsymbol{c}}$, say (m_1, w_1) and (m_2, w_2). We show that $m_1 \neq m_2$ by considering 2 cases. (Case 1): Suppose that $w_1 = w_2$. Then $m_1 \neq m_2$ by the uniqueness of the pairs. Done. (Case 2): Suppose instead that $w_1 \neq w_2$. Since G_W is full rank, we have that $(w_1 + w_2)G_W \neq 0$. In turn, $[m_1 w_1]G = [m_2 w_2]G$ implies that $(m_1 + m_2)G_M = (w_1 + w_2)G_W \neq 0$, and thus, $m_1 \neq m_2$. Done. In summary, upon receiving $\boldsymbol{c}$, Bob finds that at least 2 messages could be Alice's message. Thus, for PMFs $P_M = \mathrm{Unif}(\mathcal{M})$ and $P_W = \mathrm{Unif}(\mathcal{W})$,

$$P_{\mathrm{error}}^{\max}(\mathcal{C}_n) \geq \mathbb{P}_{(M,W) \sim P_M P_W} \left(\widehat{M} \neq M \right)$$
$$= \sum_{\boldsymbol{c} \in \mathcal{C}_n} \mathbb{P}_{(M,W) \sim P_M P_W} \left(\widehat{M} \neq M \Big| \text{Bob RXs } \boldsymbol{c} \right) \frac{1}{|\mathcal{C}_n|}$$
$$\geq 1/2.$$

Suppose instead that G_W is not full rank. Then each $(R'n)$-bit sequence in the rowspace of G_W corresponds to multiple (i.e., redundant) keys in $\mathcal{W}$. Hence, we can eliminate this redundancy by shortening the key w from $R'n$ bits to $\mathrm{rank}(G_W)$ bits and replacing G_W with full rank matrix G_W' that has $\mathrm{rowspace}(G_W') = \mathrm{rowspace}(G_W)$ without changing the code $\mathcal{C}_n$.

REFERENCES

[1] L. H. Ozarow and A. D. Wyner, "Wire-tap channel II," *AT&T Bell Laboratories Technical Journal*, vol. 63, pp. 2135–2157, 1986.

[2] P. Wang and S. Safavi-Naini, "A model for adversarial wiretap channels," *IEEE Trans. Inf. Theory*, vol. 62, pp. 970 – 983, 2016.

[3] S. Goldwasser and S. Micali, "Probabilistic encryption," *J. Computer and System Science*, vol. 28, pp. 270–299, 1984.

[4] M. Bellare, A. Desai, E. Jokipii, and P. Rogaway, "A concrete security treatment of symmetric encryption," in *Proc. IEEE Symp. Foundations of Computer Science*, Aug 1997, pp. 394–403.

[5] M. Bellare, S. Tessaro, and A. A. Vardy, "A cryptographic treatment of the wiretap channel," in *Proc. Adv. Cryptol. (CRYPTO)*, Aug 2012, pp. 1–31.

[6] C. Wang, "On the capacity of the binary adversarial wiretap channel," in *Proc. of the 54th Annual Allerton Conference on Communications, Control and Computing*, 2016.

[7] O. Ozel and U. Sennur, "Wiretap channels: implications of the more capable condition and cyclic shift symmetry," *IEEE Trans. Inf. Theory*, vol. 59, pp. 2153 – 2164, 2013.

[8] V. Guruswami and P. Indyk, "Expander-based constructions of efficiently decodable codes," in *Proc. IEEE Symp. on Foundations of Computer Science*, 2001, p. 658–667.

[9] A. D. Wyner, "The common information of two dependent random variables," *IEEE Trans. Inf. Theory*, vol. 21, pp. 163–179, 1975.

[10] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Semantic-security capacity for wiretap channels of type II," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3863–3879, 2016.

[11] H. Mahdavifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 57, pp. 6428–6443, 2011.

[12] L. A. Bassalygo, "New upper bounds for error-correcting codes," *Problems of Information Transmission*, vol. 1, pp. 32–35, 1965.

[13] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. I," *Information and Control*, vol. 10, pp. 65–103, 1967.

[14] ——, "Lower bounds to error probability for coding on discrete memoryless channels. II," *Information and Control*, vol. 10, pp. 522–552, 1967.

[15] V. Guruswami, "List decoding of error-correcting codes," Ph.D. dissertation, Massachusetts Institute of Technology, 2001. [Online]. Available: http://hdl.handle.net/1721.1/8700

[16] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. North-Holland Publishing Company, 1977.

[17] M. Cheraghchi, F. Didier, and A. Shokrollahi, "Invertible extractors and wiretap protocols," *IEEE Trans. Inf. Theory*, vol. 58, pp. 1254–1274, 2012.

[18] R. A. Chou, "Explicit wiretap channel codes via source coding, universal hashing, and distribution approximation, when the channels' statistics are uncertain," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 117–132, 2022.

[19] V. Guruswami and A. Smith, "Optimal rate code constructions for computationally simple channels," *Journal of the ACM*, vol. 63, no. 4, pp. 1–37, 2016.

[20] S. Sharifian, F. Lin, and R. Safavi-Naini, "Hash-then-encode: a modular semantically secure wiretap code," in *Proc. Workshop on Comm. Security*, July 2018, pp. 49–63.

[21] E. Ruzomberka, H. Nikbakht, C. G. Brinton, and H. V. Poor, "On pseudolinear codes for correcting adversarial errors," in *Proc. IEEE Symp. Foundations of Computer Science*, to appear, available on arXiv.

[22] K. Hoffman and R. Kunze, *Linear Algebra*, 2nd ed. Prentice-Hall, 1971.

[23] M. Plotkin, "Binary codes with specified minimum distance," *IRE Trans. Inf. Theory*, vol. 6, pp. 445–450, 1990.

[24] A. Rudra, "Lecture notes in error correcting codes: Combinatorics, algorithms and applications," https://cse.buffalo.edu/faculty/atri/courses/coding-theory/lectures/lect16.pdf, October 2007.

[25] J. P. Schmidt, A. Siegel, and A. Srinivasan, "Chernoff-Hoeffding bounds for applications with limited independence," *SIAM Journal on Discrete Mathematics*, vol. 8, no. 2, pp. 223–250, 1995.

[26] M. Bellare and J. Rompel, "Randomness-efficient oblivious sampling," in *Proc. IEEE Symp. Symposium on Foundations of Computer Science*, Aug 1994, pp. 276–287.

[27] V. Guruswami and J. Mosheiff, "Punctured low-bias codes behave like random linear codes," in *Proc. IEEE Symp. Foundations of Computer Science*, 2022, pp. 36–45.